Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 223 (2003) 299–307
Journal of Theoretical Biology
www.elsevier.com/locate/jtbi

Structural determinants of the rate of protein folding

Bengt Nölting[a,*], Wolfram Schälike[a], Patrick Hampel[a], Florian Grundig[a], Siegfried Gantert[a], Nicole Sips[a], Wolfhard Bandlow[b], Phoebe X. Qi[c]

[a] *Prussian Private Institute of Technology at Berlin, Am Schlosspark 30, Berlin D-13187, Germany*
[b] *Institute for Genetics and Microbiology, Maria-Ward-Strasse 1a, Munich D-80638, Germany*
[c] *US Department of Agriculture, ARS, ERRC, 600 E. Mermaid Lane, Wyndmoor, PA 19038, USA*

Received 28 June 2002; received in revised form 5 December 2002; accepted 14 February 2003

Abstract

To understand the mechanism of protein folding and to assist rational design of fast-folding, non-aggregating and stable artificial enzymes, it is essential to determine the structural parameters which govern the rate constants of folding, k_f. It has been found that $-\log k_f$ is a linear function of the so-called chain topology parameter (*CTP*) within the range of $10^{-1}\,s^{-1} \leqslant k_f \leqslant 10^{8}\,s^{-1}$. The correlation between $-\log k_f$ and *CTP* is much improved than using previously published contact order (*CO*) method. It has been further suggested that short sequence separations may be preferred for the establishment of stable interactions for the design of novel artificial enzymes and the modification of slow-folding proteins with aggregating intermediates.

Keywords: Protein folding kinetics; Chain topology; Φ-value analysis; Helix–coil transition; Nucleation–condensation mechanism; Rational design of enzymes

1. Introduction

There has been much studies in the field of protein folding including protein thermostability (Backmann et al., 1998; Maes and Backmann, 1999), hydrophobic (Drablos, 1999; Garcia-Hernandez and Hernandez-Arana, 1999; Chan, 2000; Czaplewski et al., 2000), hydrophilic (Jésior, 2000), electrostatic (Åqvist, 1999) and sidechain (Galzitskaya et al., 2000) interactions. In addition to its important implications in disease-related research (Bellotti et al., 1998; Ironside, 1998; Brown et al., 1999, 2000; Gursky, 1999; Kienzl et al., 1999; Gursky and Alehkov, 2000), protein folding (Nölting, 1999b; Nölting and Andert, 2000) has been closely associated with numerous other areas, such as RNA folding energy landscapes (Chen and Dill, 2000), chemical degradation (Maleknia and Downard, 2001), cell surface sialylation (Effertz et al., 1999), design of sequences with good folding properties (Irbäck et al., 1999), evolution of structure formation (D'Alessio, 1999), protein secretion (Chambert and Petit-Glatron, 1999), high-level protein expression (Hardesty et al., 1999), chaperone-assisted folding (Kawata et al., 1999; Gutsche et al., 2000) and macromolecular crowding inside the cell (Ellis and Hartl, 1999; van den Berg et al., 2000).

It has been demonstrated convincingly (Goto and Aimoto, 1991; Fersht et al., 1992; Dill et al., 1993; Karplus and Weaver, 1994; Orengo et al., 1994; Abkevich et al., 1995; Govindarajan and Goldstein, 1995; Hamada et al., 1995; Itzhaki et al., 1995; Nölting et al., 1995; Fersht, 1995a, b; Gross, 1996; Kuwajima et al., 1996; Unger and Moult, 1996; Wolynes et al., 1996; Gruebele, 1999; Forge et al., 2000; Griko, 2000; Niggemann and Steipe, 2000) that the protein folding rate is directly related to the three-dimensional structure. Topology packing (Efimov, 1999; Grigoriev et al., 1999; Clementi et al., 2000) has emerged as an important concept in understanding the relationship between protein structure and its folding process. One of the key questions in the studies of structural contacts is the interplay between short- and long-range interactions during the folding reaction (Tanaka and Scheraga, 1975, 1977; Gromiha and Selvaraj, 1997, 1999; Goto et al., 1999). It has been shown that the folding rate constant, k_f, of a protein greatly depends on the contact order, which is a measure of complexity of the protein molecule

*Corresponding author. Tel.: +49-30-485-51-81.
E-mail address: nolting@pitb.de (B. Nölting).

0022-5193/03/$ - see front matter
doi:10.1016/S0022-5193(03)00091-2

chain topology (Doyle et al., 1997; Chan, 1998; Jackson, 1998; Plaxco et al., 1998; Alm and Baker, 1999; Baker and DeGrado, 1999; Muñoz and Eaton, 1999; Riddle et al., 1999; Baker, 2000; Grantcharova et al., 2000). Proteins with a simple native chain topology, i.e. of which the native structure is dominated by contacts of residues near in sequence (Fig. 1a), are predicted to fold much faster than those with a complicated chain topology (Fig. 1b), i.e. of which the native structure contains many contacts of residues remote in sequence.

To advance our understandings of the mechanism and extreme rate of protein folding as well as to provide a basic yet applicable strategy in the rational design of novel artificial proteins and re-engineering of those slow-folded proteins that often involve aggregating intermediates, it is important to address the question of how structural contacts build up during folding at a sub-nanometer resolution (Nölting, 1998, 1999a). We showed here that $-\log k_f$ correlates well with the so-called chain topology parameter, *CTP*. The definition of *CTP* is similar to the contact order used previously to describe the complexity of the chain topology of the protein molecule, but yields much improved results. Without the need of splitting the data into subsets, e.g. α-helix or sheet, the relationship of $-\log k_f \sim CTP$ is valid over a range of rate constants of $10^{-1}\,\mathrm{s}^{-1} \leqslant k_f \leqslant 10^{8}\,\mathrm{s}^{-1}$ with a correlation coefficient of up to ≈ 0.87. Using this finding, very fast rate constants may be predicted several orders of magnitude better than with previous methods.

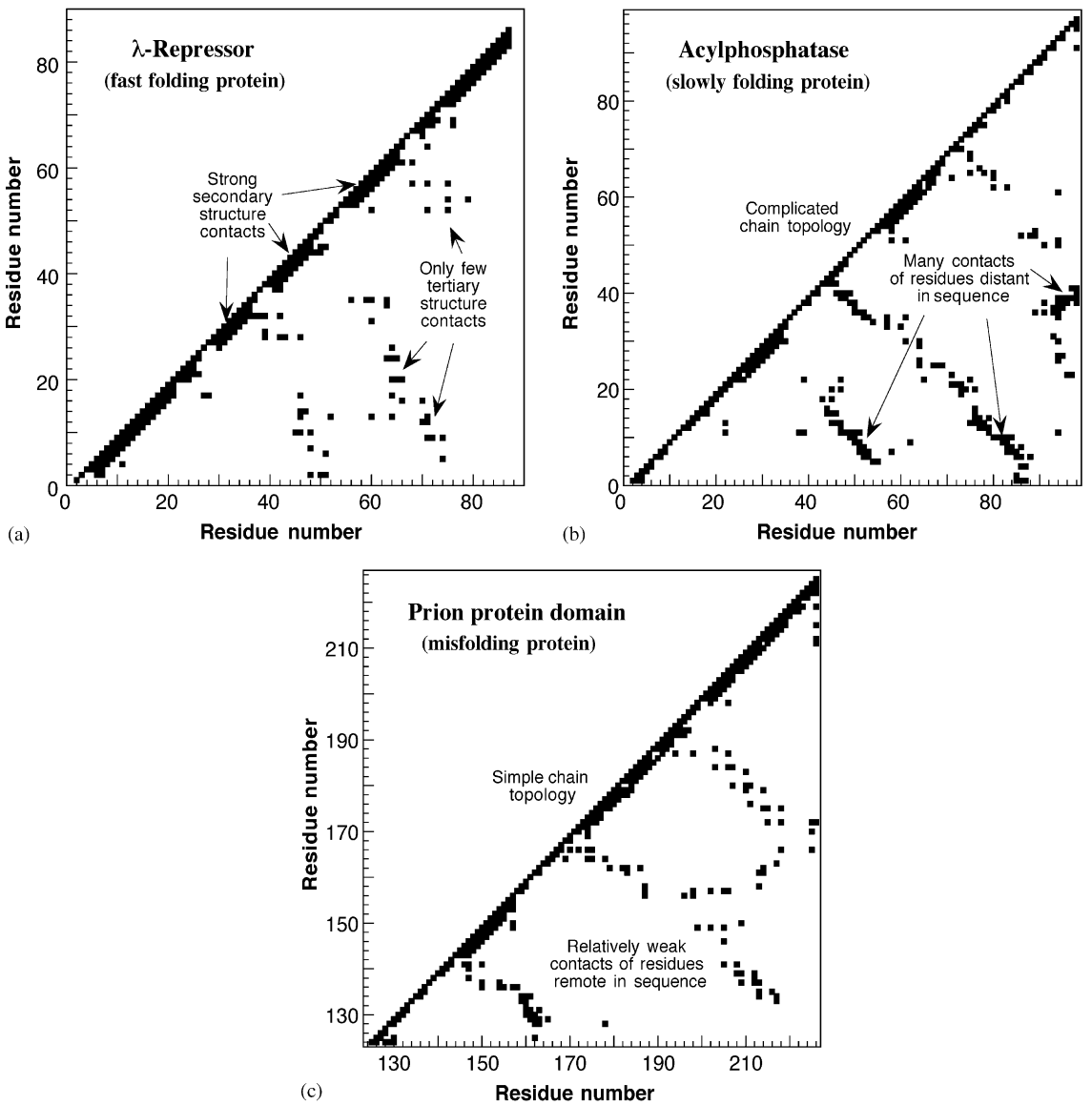

Fig. 1. (a) Inter-residue contact map of a rapid-folding protein, λ-repressor. (b) Inter-residue contact map of a slow-folding protein, acylphosphatase. (c) Inter-residue contact map of the prion protein domain PrP (121–231).

2. Materials and methods

2.1. Calculating measures of chain topology

Inter-residue contacts were calculated at a cut-off distance of 4 Å unless stated otherwise. With the exception of some data in Fig. 3, no contacts of hydrogen atoms were included in the calculations. The following 22 protein coordinates were taken from the Brookhaven National Laboratory Protein Data Bank (Abola et al., 1997): bovine acyl-coenzyme A binding protein (2ABD), mouse prion protein domain PrP(121–231) (1AG2) (only used in Fig. 1c; available only for residues 124–226), streptomyces α-amylase inhibitor tendamistat (2AIT), horse hydrolase acylphosphatase (1APS), barley chymotrypsin inhibitor 2 (CI2) (1COA), *Bacillus subtilis* major cold shock protein (CSPB) (1CSP), human fk506 binding protein (FKBP12) (1FKB), human cell adhesion protein fragment of fibronectin encompassing type-III (1FNF), Engrailed Homeodomain from *Drosophila melanogaster* (1HDD), histidine-containing phosphocarrier protein from *Escherichia coli* (HPR) (1HDN), bacteriophage λ-repressor (1LMB), thermostable variant of bacteriophage λ-repressor (1LMB) with computer-aided replacement G46A/G48A; major cold shock protein *E. coli* (CSPA) (1MJC), human SH3 domain fyn (1NYF), activation domain from porcine procarboxypeptidase B (1PBA), β-hairpin (41–56 peptide) from streptococcus immunoglobulin binding protein G (1PGB), SH3 domain human phosphatidylinositol 3-kinase (1PKS), B1 domain from *Peptostreptococcus magnus* immunoglobulin L chain binding protein (2PTL), chicken SH3 domain α-spectrin (1SHG), chicken SH3 domain src tyrosine kinase transforming protein (1SRL), third fibronectin type III repeat human tenascin (1TEN), human ribonucleoprotein U1A (1URN). Coordinates of the 10-residue helical polyalanine peptide were calculated with the program FoldIt (Jésior et al., 1994).

2.2. Folding rate constants

Eighteen rate constants (k_f) were obtained from reference (Jackson, 1998). The k_f of the 16-residue β-hairpin was chosen as previously described in Muñoz and Eaton (1999). The k_f of the 10-residue helical polyalanine peptide was estimated using data in Williams et al. (1996), Gruebele (1999), Zhou and Karplus (1999), Nölting (1999b). The k_f values for the thermostable variant of λ-repressor and for the Engrailed Homeodomain, $\approx 50{,}000\,s^{-1}$, and $37{,}000\,s^{-1}$ are from Burton et al. (1996, 1997), and Mayor et al. (2000), respectively. k_f is in units of s^{-1} throughout this article.

3. Results and discussion

Protein folding is a surprisingly fast and extremely efficient process (Fersht et al., 1992; Fersht, 1995a, b, 1998, 2000; Nölting et al., 1995; Nölting, 1999b; Nölting and Andert, 2000; Benitez-Cardoza et al., 2001; Bu et al., 2001; Ferguson et al., 2001; Galzitskaya et al., 2001; Navea et al., 2001; Ozkan et al., 2001; Rami and Udgaonkar, 2001; Roumestand et al., 2001; Volk, 2001; Vu et al., 2001; Galzitskaya et al., 2002; Kaushik et al., 2002; Yang et al., 2002).

3.1. Correlation of $-\log k_f$ with CTP

It was found that $-\log k_f$ correlates well with the so-called chain topology parameter, *CTP*:

$$-\log k_f \sim CTP, \quad CTP = \frac{1}{LN}\sum \Delta S_{i,j}^2, \tag{1}$$

where k_f is the rate constant of folding, L is the number of residues of the protein (chain length), N is the number of inter-residue contacts in the protein molecule, $\Delta S_{i,j}$ is the separation in sequence between the contacting residues number i and j, and "$\sim$" indicates a linear correlation.

The reasoning for the Eq. (1) started with the observation of a curvature of *CO* versus $-\log k_f$. It was tested which power of $\Delta S_{i,j}$ fits the best the rates $-\log k_f$. The result was close to 2 and then rounded to 2 for reasons of simplicity. Previous studies could not motivate any other exponent than 1. This may be because of smaller data sets and especially because of the absence of some peptides in the data set. The conclusion from the observation of this relation (Eq. (1)) is that the distance separation of contacts contributes nonlinearly to the free energy difference between unfolded and transition state, $\Delta G_{U-TS} = RT \ln(hk_f/(k_B T))$, where $k_B = 1.3807 \times 10^{-23}\,J\,K^{-1}$ is the Boltzmann constant, $h = 6.6261 \times 10^{-34}\,J\,s$ is the Planck constant, T is the absolute temperature, and $R = 8.3145\,J\,mol^{-1}\,K^{-1}$ is the molar gas constant (see e.g. Fersht, 1998; Nölting, 1999b).

For a set of the 20 proteins plus a 16-residue β-hairpin (Muñoz and Eaton, 1999) and a 10-residue helical polyalanine peptide, the correlation coefficient, R, is 0.86 at a cut-off distance of 4 Å for the calculation of the inter-residue contacts of (Fig. 2a). Within the range of $10^{-1}\,s^{-1} \leqslant k_f \leqslant 10^{8}\,s^{-1}$, the prediction of k_f is up to a million times better than the prediction of k_f using previous methods, e.g. the contact order, *CO* (Fig. 2b; Baker, 2000).

Fig. 2b shows that the contact order, *CO*, works particularly poorly for the two peptides. However, for the smaller data set of previous studies without peptides, the difference between the correlation coefficients for *CO* and *CTP* is much smaller. There have been recent

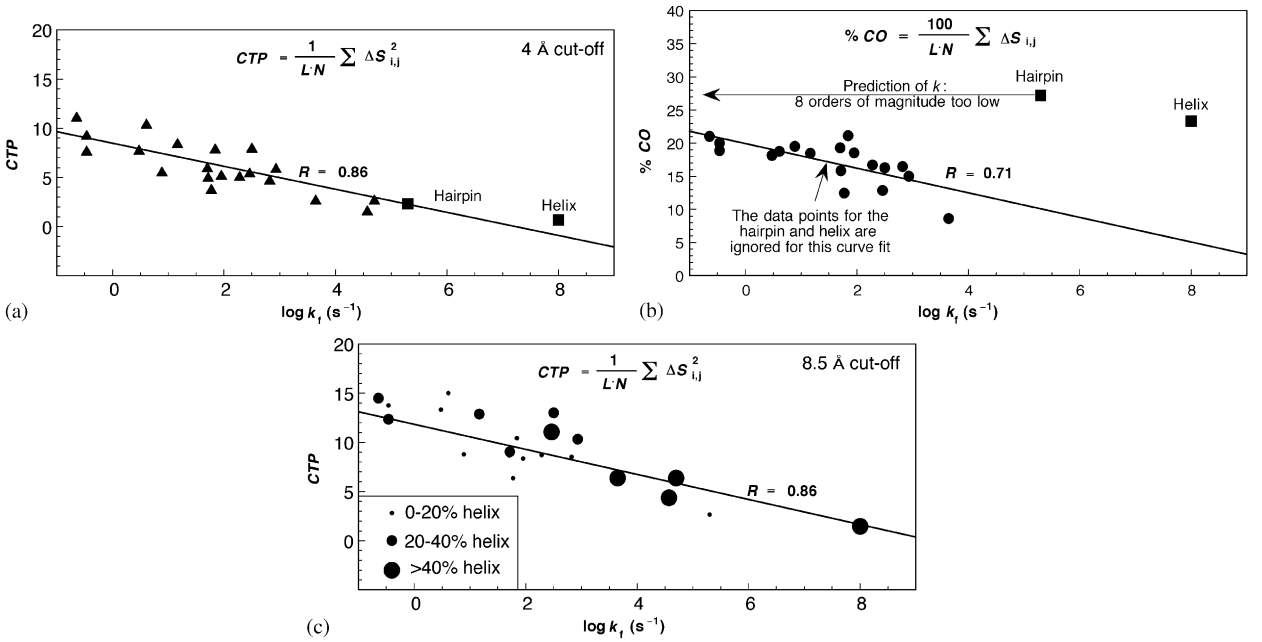

Fig. 2. (a) The calculated folding rate constants, k_f, of 20 proteins, a 16-residue β-hairpin, and a 10-residue helical polyalanine peptide the number of residues of the macromolecule using Eq. (1) for a cut-off distance of 4 Å. (b) Contact order (%*CO*) for the same set of proteins. (c) The same as in (a), but for a cut-off distance of 8.5 Å.

reports of similar empirical topology metrics related to folding rate (see Refs. in Makarov and Plaxco, 2003), but none of them work for isolated helices and hairpins as good as *CTP*.

3.2. Fit stability

For various cut-off distances from 3.5 to 8 Å, the correlation coefficient, R, for $-\log k_f \sim CTP$ is 0.80–0.87 (Fig. 3). Ignoring the contacts involving hydrogen atoms of which the positions usually are less precisely known or not fixed, e.g. due to rapid molecular motions, causes only little if any effect on R (Fig. 3). When ignoring the data points for the small peptides, the R for $-\log k_f \sim CTP$ is still 0.75–0.81 for this range of cut-off distances. The correlation thus is relatively stable regardless how the data points were selected.

There is, in general, an extensive literature suggesting a larger cut-off distance, R_c. Jernigan and Bahar (1996), for example, have shown that the Lennard–Johnes potential between β-carbons decays to zero at around $R_c = 7.5$ Å. If we add another approximately 1 Å to account for $C_\beta - C_\alpha$ distance, the reasonable R_c between α-carbons seems to be 8.5 Å. Fig. 2c shows the fit for this cut-off distance. The correlation coefficient is still 0.86. Also here the correlation is not as good for the smaller data sets without peptides of previous studies. In fact, for the smaller previous data sets without peptides, the difference between the correlation coefficients for *CO* and *CTP* may not be significant.

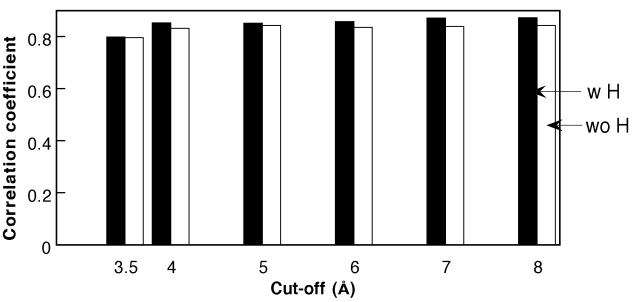

Fig. 3. Correlation coefficient for $-\log k_f \sim CTP$ with different cut-off distances for the calculation of the contacts, with and without the contacts of hydrogen atoms included as indicated.

3.3. Predictions of other folding rate constants

For a simple helical structure with very few long-range contacts, $CTP \approx 2$, and the k_f is predicted to be 10^6–$10^7\,s^{-1}$, which agrees well with the estimated diffusion limit of folding of $1\,\mu s^{-1}$ (Hagen et al., 1996) as well as the early groundbreaking measurements of helix–coil transitions (Hammes and Roberts, 1969).

3.4. Other factors affecting the folding rate constants

Clearly, there must be other factors affecting the rate of folding. For example, members of the fatty acid binding protein family have been shown to have folding rates that differ by three orders of magnitude, and yet these proteins have backbone structures that superimpose with RMSD values of 1 Å or less (Burns et al.,

1998). Intuitively, these proteins would have similar *CTP* values and similar predicted folding rates; clearly, this is not the case. Some single or double mutations that cause little if any change in *CTP* can cause significant changes in folding rate constants (Nölting et al., 1997). This suggests that other strong interactions, e.g. charge interactions, are a further important structural determinant of the folding rate. $-\mathrm{Log}\,k_f$ correlates comparably well ($R = 0.73$) with the number of the residues that belong to β-sheets. This may be explained by the larger number of long-range secondary structure contacts in sheets relative to helices. A weak correlation is also found between $-\log k_f$ and the chain length, N (not shown).

The nature of correlation between *CTP* might be questionable due to dramatic stratification of fast folders from slow folders. It is known that α-helical proteins fold faster, while β-proteins do slower. It could be possible that all of the measures of topology of the native states just reflect the secondary structure organization of proteins. To answer this question, the *CTP* values in Fig. 2c are highlighted according to the α-helix content of the corresponding proteins and peptides. One can see that the three different groups of data points, 0–20% α-helix content, 20–40% α-helix content, and 40–100% α-helix content, respectively, show a similar tendency like the whole data set. However, the number of available data points is still too small for a meaningful comparison of the correlation coefficients.

3.5. Topology of the transition state

It seems that where topology must be important is at the transition states, because one would expect that the transition states conformation topologies would bear information about corresponding free-energy barriers (Fersht et al., 1992; Fersht, 1995a, b, 1998, 2000; Oliveberg and Fersht, 1996; Oliveberg et al., 1998; Oliveberg, 2001). There are several publications that have attempted to do just that (Abkevich et al., 1994; Dokholyan et al., 2000; Vendruscolo et al., 2002). Vendruscolo et al., for example, used clustering coefficient to characterize the topological properties of the transition state conformations. They used a different measure of protein topology—node betweenness—to study relative importance of amino acids to folding kinetics.

An analysis of the structural consolidation in the transition states of proteins with various folding rate constants was performed as described in Nölting (1998, 1999a), Nölting and Andert (2000). Fig. 4 shows the results of a rapid folding protein, Arc repressor; a protein with a moderate folding rate constant, src SH3; a slow folding protein, FKBP12; and a very slow folding protein, acylphosphatase. The available sets of mutants were not sufficient to make conclusive predictions for all contacts; however, it is apparent that the complexity of the topology of the transition state undergoes significant changes as a function of the folding rate constant. The faster the rate of folding the simpler the topology of contact formation is in the transition state.

3.6. Why is there such a relationship with the native structure?

It is strange that one would expect correlation of the native protein structure with the kinetic properties of the proteins. Native structure has little information about kinetic barriers—look, for example, at various mutants that have similar structure albeit drastically different folding rates. Of course, the argument that one may provide is that nature may have selected those sequences that have specific folding rates (see e.g. discussion in Dokholyan and Shakhnovich, 2001). Another possibility is that some protein transition states are close to their native states—CI2, for example. It is a serious conceptual issue that has been underlying the whole discussion in the field, i.e. why there is such a relationship.

Φ-value analysis on a number of proteins has shown that in many cases the transition states have low Φ-values in many parts of the molecule and thus low structural consolidation as measured by the free energy of consolidation (see e.g. Nölting and Andert 2000; and references therein). Thus, protein transition states close to their native states as judged by free energy appear to be not always the case. However, even some similarity of the transition state and native structures should already influence the statistics even when the transition state is still far away from the native state. The currently available Φ-value data sets are not sufficient to judge if this is the only reason for the relationship.

It should be noted, however, that structural changes connected with low energy changes might remain undetected by Φ-value analysis. In particular, the absence of significant structural consolidation as measured by the free energy change does not exclude the possibility of a significantly correct long-range alignment at the measured positions or even in the whole molecule. The long-range structural properties are—on the other hand—the ones which matter most for *CTP* and as well for the previously used contact order. Thus the relationship of chain topology could be an indication that at least the long-range alignment of the transition state structure is close to the native structure. This hypothesis is also substantiated by the discovery of the validity of the nucleation–condensation model for folding of several proteins (see e.g. Fersht, 1995b, 1998, 2000; Nölting and Andert, 2000). According to this model, the initial folding nucleus has some degree of stability only in the presence of some correct tertiary structure interactions. It appears quite possible that in

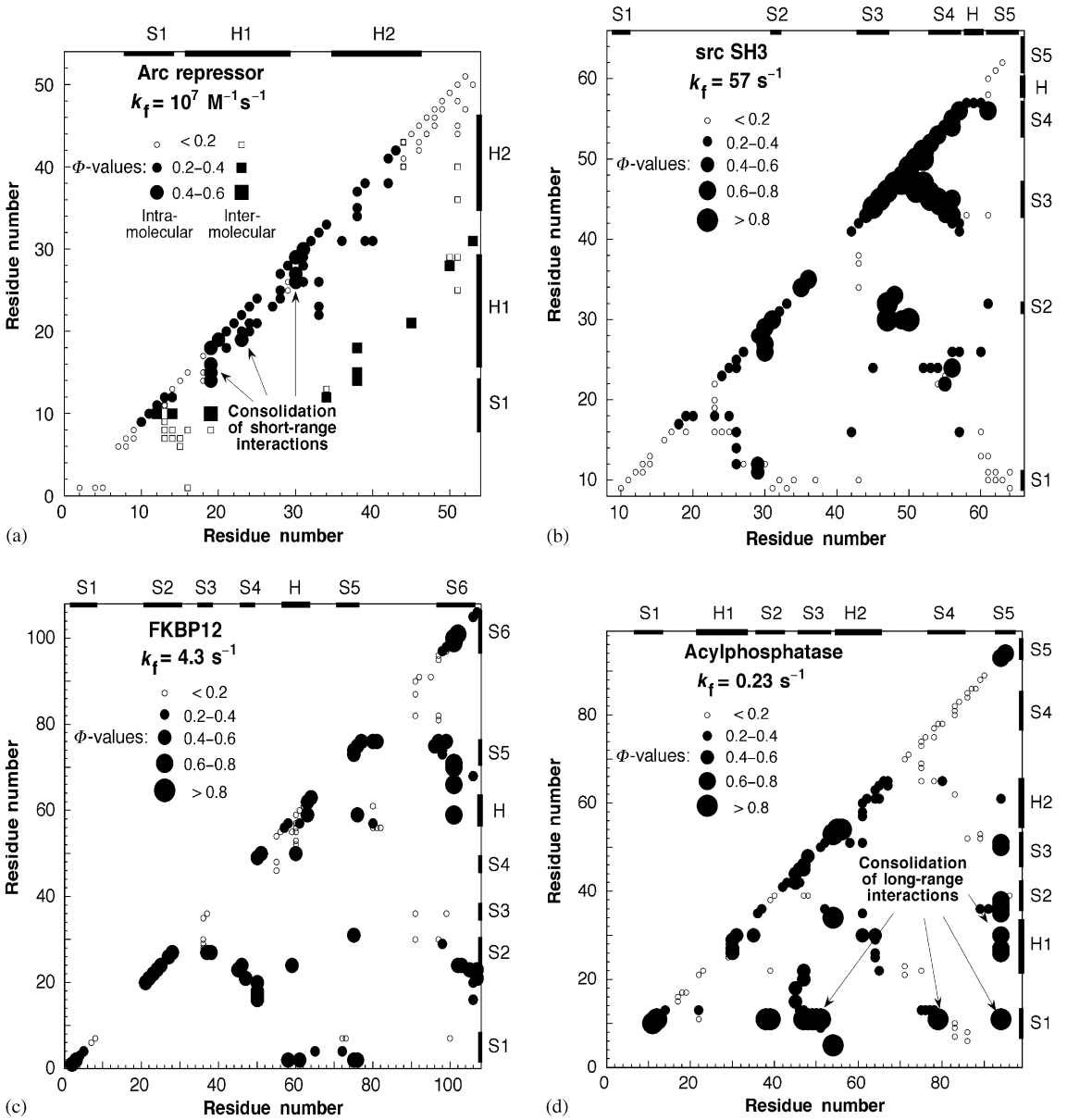

Fig. 4. Φ-values (Fersht et al., 1992; Nölting, 1999b) are correlated with inter-residue contacts; $\Phi < 0.2, 0.2 < \Phi < 0.8$, and $\Phi > 0.8$ correspond to no significant probability of structural consolidation, a medium probability of structural consolidation, and a high probability of structural consolidation, respectively. (a) Arc repressor (Milla et al., 1995): $k_f = 10^7\,M^{-1}\,s^{-1}$. (b) src SH3 domain (Riddle et al., 1999), $k_f = 57\,s^{-1}$. (c) FKBP12 (Fulton et al., 1999), $k_f = 4.3\,s^{-1}$. (d) acylphosphatase (Chiti et al., 1999): $k_f = 0.23\,s^{-1}$. Structural consolidation of the transition states was predicted as described in Nölting (1998), Nölting (1999a), Nölting and Andert, (2000).

many protein folding transition states the correct tertiary structure alignment involves more parts of the molecule than only the most consolidated residues, i.e. the nucleus. This could then further contribute to the observed relationship.

3.7. Implications in the mechanism of folding

It has been revealed in this work that a large fraction of interactions of residues close in sequence in the native structure (and already in the transition state) promotes a high rate of folding (Fig. 5). Conversely, if the structure is dominated by interactions of residues remote in sequence, folding is considerably slowed down.

Eq. (1) implies that the contribution of a contact to $-\log k_f$ is proportional to the average square of the separation of the corresponding residues in sequence. The formation of a contact between N- and C-terminus in the native structure usually takes several orders of magnitude longer than the formation of a contact of two residues half as far apart (Fig. 2a). The fit to Eq. (1) yields $-\log k_f = 7.56 - 0.895 CTP$ (Fig. 2a). As expected, this kind of behavior does not resemble the result predicted for a folding mechanism via a randomly

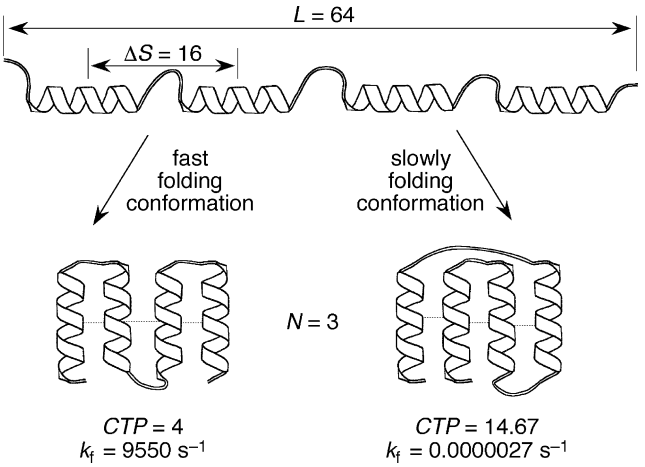

Fig. 5. Schematic illustration of two possible folding conformations of a four-helix-bundle protein with 64 residues and three contacts in the native state. The magnitudes of k_f are calculated using the relation $-\log k_f = 7.56-0.895CTP$.

sampling of all conformations where one would obtain $-\log k_f \approx L - 9$ (k_f is in units of s^{-1}). We propose that this dramatic effect of distant contacts on the folding rate constant is caused by the simultaneous build-up of secondary and tertiary structure, as emphasized in the nucleation–condensation model for folding (Itzhaki et al., 1995; Fersht, 1995a, b; Nölting et al., 1997; Shakhnovich, 1997; Nölting, 1999b). It is especially interesting to point out that the effect of the chain topology on k_f is consistent with the finding that folding nuclei of proteins with complex chain topologies have many non-local contacts. A nucleus that also possesses some degree of correct tertiary structure alignment further accelerates folding more efficiently than a nucleus that involves only secondary structure contacts.

Intriguingly, the prion protein domain, which is known to misfold, is found to have a particularly low *CTP* of 4.38 (Fig. 1c). For this protein, we speculate that the absence of sufficient long-range interactions in the folding reaction promotes misfolding.

4. Conclusions

(1) $-\mathrm{Log}\, k_f$ decreases approximately proportionally to the average square of the sequence separation in the native conformation within the range of at least $10^{-1}\,s^{-1} \leqslant k_f \leqslant 10^{8}\,s^{-1}$. This correlation between $-\log k_f$ and *CTP* is much improved than using previously published methods. It predicts the rate constants, k_f, of many small monomeric proteins within less than 1 order of magnitude compared to their experimental values.

(2) The effect of a more complex topology of the structural consolidation in the slow-folding proteins is already present in the transition states of folding.

(3) For the design of novel artificial enzymes and the modification of slow-folding proteins with aggregating intermediates, short sequence separations may be preferred for the establishment of stable interactions.

Acknowledgements

We gratefully acknowledge many highly important contributions to this study by Professor Alan R. Fersht.

References

Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I., 1994. Specific nucleus as the transition-state for protein-folding—evidence from the lattice model. Biochemistry 33, 10026–11036.

Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I., 1995. Impact of local and non-local interactions on the thermodynamics and kinetics of protein folding. J. Mol. Biol. 252, 460–471.

Abola, E.E., Sussman, J.L., Prilusky, J., Manning, N.O., 1997. Protein data bank archives of three-dimensional macromolecular structures. Methods Enzymol. 277, 556–571.

Alm, E., Baker, D., 1999. Matching theory and experiment in protein folding. Curr. Opin. Struct. Biol. 9, 189–196.

Åqvist, J., 1999. Long-range electrostatic effects on peptide folding. FEBS Lett. 457, 414–418.

Backmann, J., Schäfer, G., Wyns, L., Bönisch, H., 1998. Thermodynamics and kinetics of unfolding of the thermostable trimeric adenylate kinase from the archaeon *Sulfolobus acidocaldarius*. J. Mol. Biol. 284, 817–833.

Baker, D., 2000. A surprising simplicity to protein folding. Nature 405, 39–42.

Baker, D., DeGrado, W.F., 1999. Engineering and design. Curr. Opin. Struct. Biol. 9, 485–486.

Bellotti, V., Stoppini, M., Mangione, P., Sunde, M., Robinson, C., Asti, L., Brancaccio, D., Ferri, G., 1998. β2-microglobulin can be refolded into a native state from ex vivo amyloid fibrills. Eur. J. Biochem. 258, 61–67.

Benitez-Cardoza, C.G., Rojo-Dominguez, A., Hernandez-Arana, A., 2001. Temperature-induced denaturation and renaturation of triosephosphate isomerase from *Saccharomyces cerevisiae*: evidence of dimerization coupled to refolding of the thermally unfolded protein. Biochemistry 40, 9049–9058.

Brown, D.R., Wong, B.S., Haifiz, F., Clive, C., Haswell, S.J., Jones, I.M., 1999. Normal prion protein has an activity like that of superoxide dismutase. Biochem. J. 344, 1–5.

Brown, D.R., Hafiz, F., Glasssmith, L.L., Wong, B.S., Jones, I.M., Clive, C., Haswell, S.J., 2000. Consequences of manganese replacement of copper for prion protein function and proteinase resistance. EMBO J. 19, 1180–1186.

Bu, Z., Cook, J., Callaway, D.J., 2001. Dynamic regimes and correlated structural dynamics in native and denatured α-lactalbumin. J. Mol. Biol. 312, 865–873.

Burns, L.L., Dalessio, P.M., Ropson, I.J., 1998. Folding mechanism of three structurally similar β-sheet proteins. Proteins 33, 107–118.

Burton, R.E., Huang, G.S., Daughterty, M.A., Fullbright, P.W., Oas, T.G., 1996. Microsecond protein folding through a compact transition state. J. Mol. Biol. 263, 311–322.

Burton, R.E., Huang, G.S., Daugherty, M.A., Calderone, T.L., Oas, T.G., 1997. The energy landscape of a fast-folding protein mapped by Ala->Gly substitutions. Nat. Struct. Biol. 4, 305–310.

Chambert, R., Petit-Glatron, M.F., 1999. Anionic polymers of *Bacillus subtilis* cell wall modulate the folding rate of secreted proteins. FEMS Microbiol. Lett. 179, 43–47.

Chan, H.S., 1998. Protein folding: matching rate and locality. Nature 392, 761–763.

Chan, H.S., 2000. Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. Proteins 40, 543–571.

Chen, S.J., Dill, K.A., 2000. RNA folding energy landscapes. Proc. Natl Acad. Sci. USA 97, 646–651.

Chiti, F., Taddei, N., White, P.M., Bucciantini, M., Magherini, F., Stefani, M., Dobson, C.M., 1999. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat. Struct. Biol. 6, 1005–1009.

Clementi, C., Nymeyer, H., Onuchic, J.N., 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and en-route intermediates for protein folding? An investigation for small globular proteins. J. Mol. Biol. 298, 937–953.

Czaplewski, C., Rodziewicz-Motowidlo, S., Liwo, A., Ripoll, D.R., Wawak, R.J., Scheraga, H.A., 2000. Molecular simulation study of cooperativity in hydrophobic association. Protein Sci. 9, 1235–1245.

D'Alessio, G., 1999. The evolutionary transition from monomeric to oligomeric proteins: tools, the environment, hypotheses. Prog. Biophys. Mol. Biol. 72, 271–298.

Dill, K.A, Fiebig, K.M., Chan, H.S., 1993. Cooperativity in protein-folding kinetics. Proc. Natl Acad. Sci. USA 90, 1942–1946.

Dokholyan, N.V., Shakhnovich, E.I., 2001. Understanding hierarchical protein evolution from first principles. J. Mol. Biol. 312, 289–307.

Dokholyan, N.V., Buldyrev, S.V., Stanley, H.E., Shakhnovich, E.I., 2000. Identifying the protein folding nucleus using molecular dynamics. J. Mol. Biol. 296, 1183–1188.

Doyle, R., Simons, K., Qian, H., Baker, D., 1997. Local interactions and the optimization of protein folding. Proteins 29, 282–291.

Drablos, F., 1999. Clustering of non-polar contacts in proteins. Bioinformatics 15, 501–509.

Effertz, K., Hinderlich, S., Reutter, W., 1999. Selective loss of either the epimerase or kinase activity of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase due to site-directed mutagenesis based on sequence alignments. J. Biol. Chem. 274, 28771–28778.

Efimov, A.V., 1999. Complementary packing of α-helices in proteins. FEBS Lett. 463, 3–6.

Ellis, R.J., Hartl, F.U., 1999. Principles of protein folding in the cellular environment. Curr. Opin. Struct. Biol. 9, 102–110.

Ferguson, N., Li, W., Capaldi, A.P., Kleanthous, C., Radford, S.E., 2001. Using chimeric immunity proteins to explore the energy landscape for α-helical protein folding. J. Mol. Biol. 307, 393–405.

Fersht, A.R., 1995a. Characterizing transition states in protein folding: an essential step in the puzzle. Curr. Opin. Struct. Biol. 5, 79–84.

Fersht, A.R., 1995b. Optimization of rates of protein folding: the nucleation–condensation mechanism and its implications. Proc. Natl Acad. Sci. USA 92, 10869–10873.

Fersht, A.R., 1998. Structure and Mechanism in Protein Science. Freeman, New York.

Fersht, A.R., 2000. Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. Proc. Natl Acad. USA 97, 1525–1529.

Fersht, A.R., Matouschek, A., Serrano, L., 1992. The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. J. Mol. Biol. 224, 771–782.

Forge, V., Hoshino, M., Kuwata, K., Arai, M., Kuwajima, K., Batt, C.A., Goto, Y., 2000. Is folding of β-lactoglobulin non-hierarchic? Intermediate with native-like β-sheet and non-native α-helix. J. Mol. Biol. 296, 1039–1051.

Fulton, K.F., Main, E.R., Daggett, V., Jackson, S.E., 1999. Mapping the interactions present in the transition state for unfolding/folding of FKBP12. J. Mol. Biol. 291, 445–461.

Galzitskaya, O.V., Surin, A.K., Nakamura, H., 2000. Optimal region of average side-chin entropy for fast protein folding. Protein Sci. 9, 580–586.

Galzitskaya, O.V., Ivankov, D.N., Finkelstein, A.V., 2001. Folding nuclei in proteins. Mol. Biol. 35, 605–613.

Galzitskaya, O.V., Higo, J., Finkelstein, A.V., 2002. α-helix and β-hairpin folding from experiment, analytical theory and molecular dynamics simulations. Curr. Protein Pept. Sci. 3, 191–200.

Garcia-Hernandez, E., Hernandez-Arana, A., 1999. Structural basis of lectin–carbohydrate affinities: comparison with protein-folding energetics. Protein Sci. 8, 1075–1086.

Goto, Y., Aimoto, S., 1991. Anion and pH-dependent conformational transition of an amphiphilic polypeptide. J. Mol. Biol. 218, 387–396.

Goto, Y., Hoshino, M., Kuwata, K., Batt, C.A., 1999. Folding of β-lactoglobulin, a case of the inconsistency of local and non-local interactions. In: Kuwajima, K., Arai, M. (Eds.), Old and New Views of Protein Folding. Elsevier, Amsterdam, pp. 3–11.

Govindarajan, S., Goldstein, R.A., 1995. Optimal local propensities for model proteins. Proteins 95, 413–418.

Grantcharova, V.P., Riddle, D.S., Baker, D., 2000. Long-range order in the src SH3 folding transition state. Proc. Natl Acad. Sci. USA 97, 7084–7089.

Grigoriev, I.V., Mironov, A.A., Rakhmaninova, A.B., 1999. Refinement of helix boundaries in α-helical globular proteins. Mol. Biol. 33, 206–214.

Griko, Y.V., 2000. Energetic basis of structure stability in the molten globule state—α-lactalbumin. J. Mol. Biol. 297, 1259–1268.

Gromiha, M.M., Selvaraj, S., 1997. Influence of medium and long range interactions in different structural classes of globular proteins. J. Biol. Phys. 23, 151–162.

Gromiha, M.M., Selvaraj, S., 1999. Importance of long-range interactions in protein folding. Biophys. Chem. 77, 49–68.

Gross, M., 1996. Linguistic analysis of protein folding. FEBS Lett. 390, 249–252.

Gruebele, M., 1999. The fast protein folding problem. Annu. Rev. Phys. Chem. 50, 485–516.

Gursky, O., 1999. Probing the conformation of a human apolipoprotein C-1 by amino acid substitutions and trimethylamine-N-oxide. Protein Sci. 8, 2055–2064.

Gursky, O., Aleshkov, S., 2000. Temperature-dependent β-sheet formation in β-amyloid Aβ1–40 peptide in water: Uncoupling β-structure folding from aggregation. Biochim. Biophys. Acta 1476, 93–102.

Gutsche, I., Holzinger, J., Rossle, M., Heumann, H., Baumeister, W., May, R.P., 2000. Conformational rearrangements of an archaeal chaperonin upon ATPase cycling. Curr. Biol. 10, 405–408.

Hagen, S.J., Hofrichter, J., Szabo, A., Eaton, W.A., 1996. Diffusion-limited contact formation in unfolded cytochrome-c—estimating the maximum rate of protein-folding. Proc. Natl Acad. Sci. USA 93, 11615–11617.

Hamada, D., Kuroda, Y., Tanaka, T., Goto, Y., 1995. High helical propensity of the peptide fragments derived from β-lactoglobulin, a predominantly β-sheet protein. J. Mol. Biol. 254, 737–746.

Hammes, G.G., Roberts, P.B., 1969. Dynamics of helix–coil transition in poly L-ornithine. J. Am. Chem. Soc. 91, 1812–1816.

Hardesty, B., Tsalkova, T., Kramer, G., 1999. Co-translational folding. Curr. Opin. Struct. Biol. 9, 111–114.

Irbäck, A., Peterson, C., Potthast, F., Sandelin, E., 1999. Design of sequences with good folding properties in coarse-grained protein models. Struct. Fold Des. 7, 347–360.

Ironside, J.W., 1998. Prion diseases in man. J. Pathol. 186, 227–234.
Itzhaki, L.S., Otzen, D.E., Fersht, A.R., 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering methods: evidence for a nucleation–condensation mechanism for protein folding. J. Mol. Biol. 254, 260–288.
Jackson, S.E., 1998. How do small single-domain proteins fold? Fold. Des. 3, R81–R91.
Jernigan, R.L., Bahar, I., 1996. Structure-derived potentials and protein simulations. Curr. Opin. Struct. Biol. 6, 195–209.
Jesior, J.C., 2000. Hydrophilic frameworks in proteins? J. Protein Chem. 19, 93–103.
Jesior, J.C., Filhol, A., Tranqui, D., 1994. Foldit (light)—an interactive program for macintosh computers to analyze and display protein data bank coordinate files. J. Appl. Cryst. 27, 1075.
Karplus, M., Weaver, D.L., 1994. Protein folding dynamics: the diffusion–collision model and experimental data. Protein Sci. 3, 650–668.
Kaushik, J.K., Ogasahara, K., Yutani, K., 2002. The unusually slow relaxation kinetics of the folding–unfolding of pyrolidone carboxyl peptidase from a hyperthermophile *Pyrococcus furiosus*. J. Mol. Biol. 316, 991–1003.
Kawata, Y., Kawagoe, M., Hongo, K., Mikuya, T., Higurashi, T., Mizobata, T., Nagai, J., 1999. Functional communications between the apical and equatorial domains of GroEL through the intermediate domain. Biochemistry 38, 15731–15740.
Kienzl, E., Jellinger, K., Stachelberger, H., Linert, W., 1999. Iron as catalyst for oxidative stress in the pathogenesis of Parkinson's disease? Life Sci. 65, 1973–1976.
Kuwajima, K., Yamaya, H., Sugai, S., 1996. The burst-phase intermediate in the refolding of β-lactoglobulin studied by stopped-flow circular dichroism and absorption spectroscopy. J. Mol. Biol. 264, 806–822.
Maes, D., Backmann, J., 1999. Protein thermostability: a case study on triphosphate isomerase. Recent Res. Dev. Biochem. 1, 105–116.
Makarov, D.E., Plaxco, K.W., 2003. The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. Protein Sci. 12, 17–26.
Maleknia, S.D., Downard, K., 2001. Radical approaches to probe protein structure, folding, and interactions by mass spectrometry. Mass Spectrom. Rev. 20, 388–401.
Mayor, U., Johnson, C.M., Daggett, V., Fersht, A.R., 2000. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. Proc. Natl Acad. Sci. USA 97, 13518–13522.
Milla, M.E., Brown, B.M., Waldburger, C.D., Sauer, R.T., 1995. P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. Biochemistry 34, 13914–13919.
Muñoz, V., Eaton, W.A., 1999. A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc. Natl Acad. Sci. USA 96, 11311–11316.
Navea, S., de Juan, A., Tauler, R., 2001. Three-way data analysis applied to multispectroscopic monitoring of protein folding. Anal. Chim. Acta 446, 187–197.
Niggemann, M., Steipe, B., 2000. Exploring local and nonlocal interactions for protein stability by structural motiv engineering. J. Mol. Biol. 296, 181–195.
Nölting, B., 1998. Structural resolution of the folding pathway of a protein by correlation of Φ-values with inter-residue contacts. J. theor. Biol. 194, 419–428.
Nölting, B., 1999a. Analysis of the folding pathway of chymotrypsin inhibitor by correlation of Φ-values with inter-residue contacts. J. theor. Biol. 197, 113–121.
Nölting, B., 1999b. Protein Folding Kinetics. Springer, Heidelberg.
Nölting, B., Andert, K., 2000. Mechanism of Protein Folding. Proteins 41, 288–298.
Nölting, B., Golbik, R., Fersht, A.R., 1995. Submillisecond events in protein folding. Proc. Natl Acad. Sci. USA 92, 10668–10672.
Nölting, B., Golbik, R., Neira, J.L., Soler-Gonzalez, A.S., Schreiber, G., Fersht, A.R., 1997. The folding pathway of a protein at high resolution from microseconds to seconds. Proc. Natl Acad. Sci. USA 94, 826–830.
Oliveberg, M., 2001. Characterisation of the transition states for protein folding: towards a new level of mechanistic detail in protein engineering analysis. Curr Opin Struct Biol. 11, 94–100.
Oliveberg, M., Fersht, A.R., 1996. New approach to the study of transient protein conformations: the formation of a semiburied salt link in the folding pathway of barnase. Biochemistry 28, 6795–6805.
Oliveberg, M., Tan, Y.J., Silow, M., Fersht, A.R., 1998. The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. J. Mol. Biol. 277, 933–943.
Orengo, C.A., Jones, D.T., Thornton, J.M., 1994. Protein superfamilies and domain superfolds. Nature 372, 631–634.
Ozkan, S.B., Bahar, I., Dill, K.A., 2001. Transition states and the meaning of Φ-values in protein folding kinetics. Nat. Struct. Biol. 8, 765–769.
Plaxco, K.W., Simons, K.T., Baker, D., 1998. Contact order, transition state placement and the refolding rates of single-domain proteins. J. Mol. Biol. 277, 985–994.
Rami, B.R., Udgaonkar, J.B., 2001. pH-jump-induced folding and unfolding studies of barstar: evidence for multiple folding and unfolding pathways. Biochemistry 40, 15267–15279.
Riddle, D.S., Grantcharova, V.P., Santiago, J.V., Alm, E., Ruczinski, I., Baker, D., 1999. Experiment and theory highlight role of native state topology in SH3 folding. Nat. Struct. Biol. 6, 1016–1024.
Roumestand, C., Boyer, M., Guignard, L., Barthe, P., Royer, C.A., 2001. Characterization of the folding and unfolding reactions of a small β-barrel protein of novel topology, the MTCP1 oncogene product P13. J. Mol. Biol. 312, 247–259.
Shakhnovich, E.I., 1997. Theoretical studies of protein folding thermodynamics and kinetics. Curr. Opin. Struct. Biol. 7, 29–40.
Tanaka, S., Scheraga, H.A., 1975. Model of protein folding: inclusion of short-, medium-, and long-range interactions. Proc. Natl Acad. Sci. USA 72, 3802–3806.
Tanaka, S., Scheraga, H.A., 1977. Hypothesis about the mechanism of protein folding. Macromolecules 10, 291–304.
Unger, R., Moult, J., 1996. Local interactions dominate folding in a simple protein model. J. Mol. Biol. 259, 988–994.
van den Berg, B., Wain, R., Dobson, C.M., Ellis, R.J., 2000. Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell. EMBO J. 19, 3870–3875.
Vendruscolo, M., Dokholyan, N.V., Paci, E., Karplus, M., 2002. A small-world view of the amino acids that play a key role in protein folding. Phys. Rev. E 65, 061910.
Volk, M., 2001. Fast initiation of peptide and protein folding processes. Eur. J. Org. Chem. 14, 2605–2621.
Vu, D.M., Reid, K.L., Rodriguez, H.M., Gregoret, L.M., 2001. Examination of the folding of *E. coli* CspA through tryptophan substitutions. Protein Sci. 10, 2028–2036.
Williams, S., Causgrove, T.P., Gilmanshin, R., Fang, K.S., Callender, R.H., Woodruff, W.H., Dyer, R.B., 1996. Fast events in protein folding: helix melting and formation in a small peptide. Biochemistry 35, 691–697.
Wolynes, P.G., Luthey-Schulten, Z., Onuchic, J.N., 1996. Fast folding experiments and the topography of protein-folding energy landscapes. Chem. Biol. 3, 425–432.
Yang, J.J., Yang, H.D., Ye, Y.M., Hopkins, H., Hastings, G., 2002. Temperature-induced formation of a non-native intermediate state of the all β-sheet protein CD2. Cell Biochem. Biophys. 36, 1–18.
Zhou, Y., Karplus, M., 1999. Interpreting the folding kinetics of helical proteins. Nature 401, 400–403.